

RESEARCH ARTICLE

Population genomic SNPs from epigenetic RADs: Gaining genetic and epigenetic data from a single established next-generation sequencing approach

Marco Crotti¹  | Colin E. Adams^{1,2}  | Kathryn R. Elmer¹ 

¹Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, UK

²Scottish Centre for Ecology and the Natural Environment, University of Glasgow, Rowardennan, UK

Correspondence

Kathryn R. Elmer

Email: Kathryn.Elmer@glasgow.ac.uk

Funding information

University of Glasgow College of Medical, Veterinary and Life Sciences

Handling Editor: Oscar Gaggiotti**Abstract**

1. Epigenetics is increasingly recognized as an important molecular mechanism underlying phenotypic variation. To study DNA methylation in ecological and evolutionary contexts, epiRADseq is a cost-effective next-generation sequencing (NGS) technique based on reduced representation sequencing of genomic regions surrounding non-/methylated sites. EpiRADseq for genome-wide methylation abundance and ddRADseq for genome-wide single-nucleotide polymorphism (SNP) genotyping follow very similar library and sequencing protocols, but to date these two types of dataset have been handled separately. Here we test the performance of using epiRADseq data to generate SNPs for population genomic analyses.
2. We tested the robustness of using epiRADseq data for population genomics with two independent datasets: a newly generated single-end dataset for the European whitefish *Coregonus lavaretus*, and a re-analysis of publicly available, previously published paired-end data on corals. Using standard bioinformatic pipelines with a reference genome and without (i.e. de novo catalogue loci), we compared the number of SNPs retained, population genetic summary statistics and population genetic structure between data drawn from ddRADseq and epiRADseq library preparations.
3. We found that SNPs drawn from epiRADseq are similar in number to those drawn from ddRADseq, with 55%–83% of SNPs being identified by both methods. Genotyping error rate was <5% in both approaches. EpiRADseq-specific allele dropout was low (~1%). For summary statistics, such as heterozygosity and nucleotide diversity, there is a strong correlation between methods (Spearman's $\rho > 0.88$). Furthermore, identical patterns of population genetic structure were recovered using SNPs from epiRADseq and ddRADseq approaches.
4. We show that SNPs obtained from epiRADseq are highly similar to those from ddRADseq and are equivalent for estimating genetic diversity and population structure. This finding is particularly relevant to researchers interested in genetics and epigenetics on the same individuals because using a single epigenomic

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

approach to generate two datasets greatly reduces the time and financial costs compared to using these techniques separately. It also efficiently enables correction of epigenetic estimates with population genetic data. Many studies will benefit from a combinatorial approach with genetic and epigenetic markers and this demonstrates a single, efficient method to do so.

KEYWORDS

DNA methylation, genomics, molecular ecology, RADseq, single-nucleotide polymorphism

1 | INTRODUCTION

The study of epigenetic processes, which cause changes in gene expression without nucleotide mutation of the underlying genome sequence, in an ecological and evolutionary framework has seen an increased interest in recent years and is providing a new complexity in the genotype–phenotype map (Bossdorf, Richards, & Pigliucci, 2007; Feil & Fraga, 2012; Hu & Barrett, 2017). The best understood epigenetic mechanism in eukaryotes is DNA methylation, which involves the addition of a methyl group to cytosine, and in eukaryotes it occurs mainly in CpG dinucleotides (Metzger & Schulte, 2016). Ecological epigenetics aims to understand how DNA methylation associates with patterns of population variation and influences phenotypic diversity, local adaptation and plasticity in natural populations (Bossdorf et al., 2007; Hu & Barrett, 2017). Until recently, epigenetic research in wild populations was conducted mainly using methylation-sensitive amplification length polymorphisms, since they are cost-effective, easily applied to non-model organisms and not computationally demanding (Schrey et al., 2013). However, they have several shortcomings (see review by Schrey et al., 2013), the greatest of which is that they screen anonymous loci that then cannot be genome-referenced nor compared across studies. Recently, the field has been invigorated by new methods that take advantage of so-called next-generation sequencing (NGS) technology. One example is bisulfite sequencing, which comes in a number of variations (whole-genome, reduced representation and target sequencing of specific gene regions) and provides high resolution information about the methylation landscape (Metzger & Schulte, 2016). Despite the benefits, this technique is expensive, can result in excessive DNA degradation and usually requires a related reference genome for the species of interest, something that is still lacking for most non-model organisms (Leontiou et al., 2015; Metzger & Schulte, 2016), although some protocols work without reference genome (Klughammer et al., 2015).

EpiRADseq is a recently developed reduced representation approach (Schield et al., 2016) to study DNA methylation variation in individuals. It is based on the established ddRADseq protocol (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and involves the digestion of the genome using two restriction enzymes, with one enzyme being methylation-sensitive. A methylated locus will not be cut by the methylation-sensitive enzyme, will not be enriched by polymerase chain reaction (PCR) and thus no sequencing read is obtained

in the data. If a locus is unmethylated, it will be cut in the same way as ddRADseq and therefore enriched by PCR and sequenced. Consequently, the number of overall reads for a locus is proportional to the level of (non-)methylation and differences in the methylation level between groups can be determined (Schield et al., 2016). The advantages of this technique resemble those of genomic reduced representation approach such as ddRADseq: the possibility of sampling genome-wide, the high degree of customization in terms of the number of loci and coverage obtained, no requirement for a reference genome, the ability to map loci against a reference genome (if available) to determine to which genomic region they correspond (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Schield et al., 2016) and lower library preparation cost per individual (we estimate library preparation for 100 individuals to be £640 for epiRADseq and £2,500 RRBS, including the cost of reagents and excluding sequencing costs). However, caution is warranted because the technique is based on the recognition of restriction cut sites, so it is best suited to closely related populations or individuals, as different genetic backgrounds will impact the recognition of restriction cut sites (Schield et al., 2016).

Combining genetic and epigenetic analyses in the same study is, to date, underleveraged but particularly valuable for providing insight into the relationship between that variation and downstream effects of interest, such as phenotypic diversity (Hu & Barrett, 2017). To infer methylation with epiRADseq and genomic polymorphism (single-nucleotide polymorphisms [SNPs]) using separate NGS libraries for the same set of individuals is expensive, inefficient and time consuming but is the approach that has been used to date (e.g. Dimond, Gamblewood, & Roberts, 2017). A benefit of independent genomic (ddRADseq) and epigenomic (epiRADseq) evaluation is that the presence/absence of an epiRADseq locus can be confirmed by the ddRADseq genotype to validate the methylation state of a locus (see Dimond et al., 2017; Schield et al., 2016). However, a combined molecular approach that allows for DNA methylation and genetic analyses using a single dataset, if robust, would increase the efficiency and scope of possible research questions in this area, and therefore be of considerable value to ecology and evolution studies.

Because epiRADseq is similar in molecular methodology to ddRADseq, here we test whether the SNPs recovered by epiRADseq loci can be used for population genomics. If such SNPs can be reliably extracted, then epigenomic and population genomic analyses can be conducted efficiently on the same samples using the

same molecular technique, from DNA extraction through to library preparation and sequencing. We tested this with two independent examples from natural animal populations for which epiRADseq and ddRADseq data are available from the same individuals: a previously published dataset (Dimond et al., 2017) from a marine invertebrate, the corals of the genus *Porites* (genome size between 420 Mb and 1.14 Gb) with no reference genome available; and a newly generated dataset from a vertebrate, the freshwater European whitefish *Coregonus lavaretus* (genome size 3.3 Gb) for which genome scaffolds were available. We ran analyses in parallel on epiRADseq and ddRADseq data to compare the number of SNPs retained, allele dropout (ADO), summary statistics and population genetic structure. We conclude that epiRADseq data are likely appropriate for population genomics after rigorous assessment and suggest a bioinformatic pipeline for extracting SNPs. We offer recommendations for robust SNP calling from epiRADseq data that should be explored before use in empirical studies and caution that more biological examples are needed.

2 | MATERIALS AND METHODS

2.1 | Coral data source

Dimond et al. (2017) assessed population genetics and epigenetics of three morphospecies of coral *Porites* spp. EpiRADseq was used for differential methylation analysis and ddRADseq to estimate population structure and to correct for the genotypic bias of epiRADseq in the methylation analysis, as a missing locus could either mean a lack of site due to mutation (a genetic factor) or due to methylation (an epigenetic factor; Schield et al., 2016). They excluded from the dataset epiRADseq loci that had zero reads in the ddRADseq dataset. However, they did not test the possibility of using epiRADseq to call SNPs.

2.2 | Coral data processing

The raw ddRADseq and epiRADseq reads from Dimond et al. (2017) were downloaded from http://owl.fish.washington.edu/nightingales/Porites_spp/. DNA for their ddRADseq library was digested using the enzymes *Pst*I (5'-CTGCAG-3' recognition site) and *Msp*I (5'-CCGG-3' recognition site), while the epiRADseq library was digested using the enzymes *Pst*I and *Hpa*II (5'-CCGG-3' recognition site). *Hpa*II does not cleave DNA when a 5-methyl group on the internal cytosine is present, while its isoschizomer *Msp*I cleaves DNA irrespective of methylation. Samples were pooled into 12 libraries, size-selected for fragments 415–515 bp using PippinPrep (Sage Science) and sequenced on the Illumina HiSeq400 with 100 bp paired-end reads.

Raw reads were demultiplexed with *process_radtags* in *Stacks* v.2.1 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), and trimmed to 95 bp. We then trimmed the first 5 and 3 bp with *Trimmomatic* (Bolger, Lohse, & Usadel, 2014) from the forward and reverse reads to remove the enzyme cut site, and paired-end

trimming was done with the following settings: LEADING = 20, TRAILING = 20, to remove low-quality reads, and CROP = 85, so that reads were all of the same length. Reads were mapped against the genome of the coral symbiont *Symbiodinium minutum*, provided in the Supplementary information of Dimond et al. (2017), to remove symbiont reads from the de novo assembly (following Dimond et al., 2017), using *bwa mem* v.0.7.17 (Li & Durbin, 2009). The retained coral reads were used for all further analyses.

A pseudo-reference genome of coral samples was created using the ddRADseq fastq files free of symbiont reads, so that we could determine the number of SNPs found in both the ddRADseq and epiRADseq datasets. This pseudo-genome was assembled using *Rainbow* v.2.0.4 (Chong, Ruan, & Wu, 2012) with the cluster, divide and merge functions with default parameters. *CD-Hit* v.4.7 (Fu, Niu, Zhu, Wu, & Li, 2012) was then used with the *cd-hit-est* (at a 90% identity threshold) function for further filtering. ddRADseq and epiRADseq reads were mapped against this pseudo-genome using *bwa mem* with default settings and retained if mapping quality was >20.

If a sample had fewer than 200,000 reads in either the ddRADseq or epiRADseq dataset, it was removed from both so that the datasets had the same individuals. This threshold was used to maximize the number of samples retained in the analysis. This excluded four samples, so 52 samples (26 for ddRADseq and 26 for epiRADseq) were retained. The *ref_map.pl* pipeline in *Stacks* v.2.1 (Catchen et al., 2013) was run for both ddRADseq and epiRADseq using default parameters. For this, all the samples were considered as part of the same population. The dataset was then filtered with the following parameters from the *populations* program: *-r* = 1 (no missing data allowed, same as in Dimond et al., 2017), *--min_maf* = 0.10 (alleles need to be present in at least 2.6 individuals), *--max_obs_het* = 0.6 (to remove potential paralogous sites) and *--write_single_snp*.

2.3 | Whitefish data generation

Using existing tissue samples of *Coregonus lavaretus* from four Scottish loch populations preserved in ethanol (M. Crotti, C.E. Adams, & K.R. Elmer, unpubl. data), DNA was extracted from tissues from the same individuals (fin clips for the ddRADseq and muscle tissue for the epiRADseq libraries) using the NucleoSpin Tissue kit (Macherey-Nagel) following the manufacturers' recommendations. These populations are part of a conservation translocation project and consist of two source and two translocated populations. Genomic analyses should not be influenced by tissue type, while DNA methylation is tissue-specific. Therefore, tissue selection will be important during experimental design (see Section 4.4). The protocol for ddRADseq library preparation follows Jacobs, Hughes, Robinson, Adams, and Elmer (2018). Briefly, 1 µg of genomic DNA per sample was double digested using *Pst*I-HF (5'-CATCAG-3' recognition site) and *Msp*I (New England BioLabs). Combinatorial barcoded Illumina adapters were then ligated to *Pst*I-HF and *Msp*I overhangs. Samples were size-selected

using a PippinPrep (Sage Science) at a target range of 150–300 bp fragments. For enrichment, we performed PCR amplification cycles: 30 s at 98°C, 9× (10 s 98°C, 30 s 65°C, 30 s 72°C) and 5 min 72°C. After PCR purification, the library was run on a 1.25% agarose gel stained with SYBR Safe (Life Technologies) to remove any adapter dimers and/or fragments outside the selected size range. DNA was excised manually, cleaned and quantified using Qubit Fluorometer with the dsDNA BR Assay (Life Technologies) to ensure the final library concentration of >1 ng/μl.

The protocol for the epiRADseq library was identical to the ddRADseq, except the methylation-sensitive *HpaII* was used instead of *MspI*.

The ddRADseq and epiRADseq libraries consisted of the same 43 samples each, including two technical replicates to estimate sequencing error (Mastretta-Yanes et al., 2015), and were sequenced on a single lane to 4 million reads per individual (per Jacobs et al., 2018). NGS sequencing was carried out at Glasgow Polyomics on the Illumina NextSeq 500 with 75 bp paired-end reads.

2.4 | Whitefish data processing

Processed epiRADseq and ddRADseq data were analysed separately using the same approaches. Samples with fewer than 350 K aligned reads in one dataset were excluded from both datasets. Again, this threshold was used to maximize the number of samples retained in the analysis. The filtering steps applied to the whitefish data were similar as used in the coral data, but with some modifications because the whitefish data were analysed as single-end. First, raw reads were demultiplexed with *process_radtags* and trimmed to 65 bp, with only forward reads retained. *Trimomatic* was used with following settings: HEADCROP = 5, LEADING = 20, TRAILING = 20 and MINLEN = 60. Reads were then mapped to an unpublished draft genome scaffolds of the lake whitefish *Coregonus clupeaformis* (L. Bernatchez, pers. comm.) using *bwa mem* with default settings and retained if mapping quality was >20 with *samtools* v.1.7 (Li et al., 2009). In *Stacks*, the *ref_map.pl* script was used to assemble reads into *Stacks* and call loci, and the *population* module was used to call SNPs.

To assess the sensitivity of SNP calling to missing data for epiRADseq data, we created three different datasets for both the ddRADseq and epiRADseq reads, which varied according to the proportion of individuals per population the locus had to be in to be retained (*-r* parameter): 0.667, 0.75 or 1. The other filtering parameters were kept constant: *-p* = 2, *--max_obs_het* = 0.6, *--min_maf* = 0.10 and *--write_single_snp*. The three datasets are hereafter referred to as the *-r* 67, *-r* 75 and *-r* 100 datasets. This assessment was done only for the whitefish data, as with the coral data we focused on comparing our results to the original paper (Dimond et al., 2017).

2.5 | Population genetic analysis

For both the whitefish and coral data, we recorded the total number of SNPs retained by ddRADseq and epiRADseq datasets. Summary

statistics of genetic diversity (expected heterozygosity, observed heterozygosity and nucleotide diversity) per locus calculated by the *population* module of *Stacks* for the ddRADseq and epiRADseq datasets were compared using Spearman correlation in R v.3.5.1 (R Core Team, 2018).

To compare estimates of population structure between the ddRADseq and epiRADseq datasets, we used the R package *ADEGENET* v.2.1.1 (Jombart, 2008) for a discriminant analysis of principal components (DAPC; Jombart, Devillard, & Balloux, 2010), which uses *k*-means clustering and the Bayesian information criterion to identify the most likely number of genetic clusters in the dataset. The *xvalDAPC* function was used to determine the number of PCs to be retained by the DAPC analysis. The divergence estimate between the inferred clusters was calculated using Weir and Cockerham F_{ST} (Weir & Cockerham, 1984) implemented in the R package *HIERFSTAT* v.0.04 (Goudet, 2005). For the coral analysis, we additionally ran the DAPC on the set of SNPs used by Dimond et al. (2017), which they made available in the supplementary information of their article, to compare our results to the original study.

2.6 | Genotyping error rate

To estimate genotyping error rate for the whitefish data, we used two approaches: (a) we computed a matrix of genetic distances between individuals using the function *dist.gene* in the R package *APE* v.5.2 (Paradis & Schielp, 2019), following Dimond et al. (2017); (b) we used the R script published by Mastretta-Yanes et al. (2015), where the number of SNP mismatches is counted and calculated as the ratio over all compared loci (Recknagel, Jacobs, Herzyk, & Elmer, 2015). Replicated samples were compared at six-fold coverage. Technical replicates were not included in the coral dataset, so genotyping error was not quantified.

2.7 | Allele dropout

To measure ADO leading to incorrect assignment of homozygotes, we calculated heterozygosity at each locus for every individual in *Stacks* for the ddRADseq and epiRADseq datasets for coral and whitefish separately. For each species, we created a table of shared loci and individuals for the ddRADseq and epiRADseq datasets, where each observation (row) represented an individual and a genomic site. We then calculated how many times an individual is assigned a homozygous genotype in the epiRADseq data and a heterozygous genotype in the ddRADseq data at a locus (i.e. due to methylation at one allele, called epiRADseq-specific ADO, or due to technical artefacts). By dividing the number of observations for which we observe epiRADseq-specific ADO by the number of total observations, we obtained the average proportion of loci per individual that exhibit epiRADseq-specific ADO. We also calculated the reverse cases, where a locus is heterozygous in epiRADseq and homozygous in ddRADseq (i.e. ddRADseq-specific ADO), which we consider to be exclusively technical artefact due to library preparation, sequencing and/or data coverage.

3 | RESULTS

3.1 | Coral data filtering

The 30 ddRADseq samples had a total of 213 M raw reads and the 30 epiRADseq samples a total of 156 M raw reads (Table 1). After filtering with *Trimmomatic*, the ddRADseq samples retained 205 M reads, and the epiRADseq samples retained 149 M reads. Mapping against the pseudo-genome created from the ddRADseq reads (418,401 contigs) retained 142 M reads for the ddRADseq and 102 M reads for the epiRADseq samples.

The *Stacks* pipeline generated a catalogue of 285,987 loci for the ddRADseq dataset, with a mean effective per sample coverage of 64.9×, and 164,411 loci for the epiRADseq dataset, with an effective per sample mean coverage of 75.7×. The average number of loci per individual was 58,896 for the ddRADseq and 33,843 for the epiRADseq catalogues.

3.2 | Coral data analyses

The *population* filtering generated datasets of 1,046 SNPs and 819 SNPs for ddRADseq and epiRADseq respectively (Figure 1a). The number of SNPs retained in our study is slightly lower to those used by the original study (1,113 SNPs from ddRADseq, also assessed here). By mapping reads to a reference assembly, we could calculate the number of SNPs that overlapped between the two datasets. In total 676 SNPs overlapped, which corresponds to 83% of SNPs in the epiRADseq and 65% of SNPs in the ddRADseq datasets.

Discriminant analysis of principal components analyses of the epiRADseq and ddRADseq datasets recovered the same three clusters as were inferred from the original study by Dimond et al. using ddRADseq (Figure 2). Our F_{ST} estimates between clusters ranged from 0.24 to 0.26, while the estimates of Dimond et al. were 0.19 to 0.21 (Figure 2a–c). The proportion of variation explained by the discriminant functions was similar in all three datasets (Figure 2). When comparing estimates of genetic diversity, we recovered strong Spearman's σ correlation for all three summary statistics between the ddRADseq and the epiRADseq datasets (Figure 3).

3.3 | Whitefish sequencing results and data filtering

The whitefish ddRADseq library generated a total of 524 M reads and the epiRADseq library generated 554 M reads (Table 1). After demultiplexing with *process_radtags* and filtering with *Trimmomatic*,

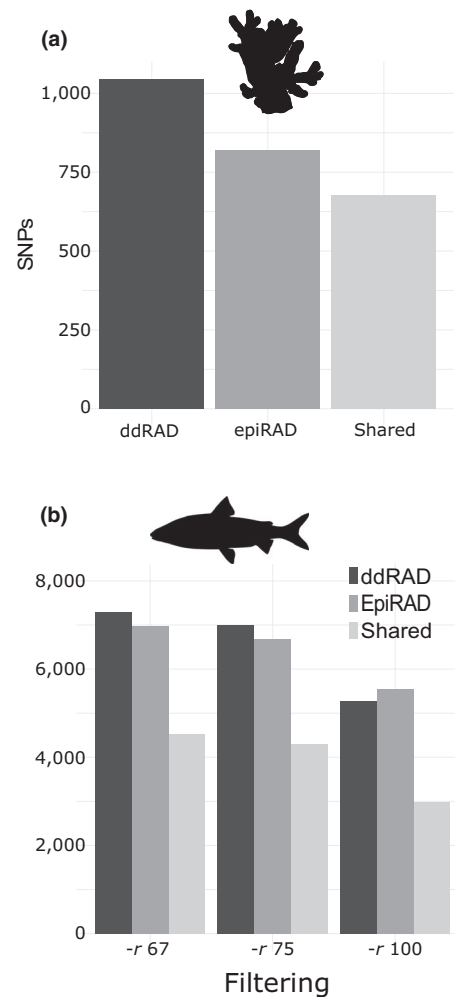


FIGURE 1 The number of single-nucleotide polymorphisms (SNPs) retained by the ddRADseq and epiRADseq datasets for (a) the coral data and (b) whitefish data. Three datasets were created for the whitefish data, differing in the percentage of individuals that must possess a particular locus for it to be included ($-r$ parameter of the *population* program from the *Stacks* pipeline)

TABLE 1 Number of samples in the libraries and number of reads retained (in millions, M) after each step. Retained reads are the number after demultiplexing and *Trimmomatic*. BAM records refer to the number of reads retained after mapping to (pseudo)reference draft genome. Catalogue loci are the total loci inferred from *Stacks*, whether variable or not

	Number of individuals	Total reads (millions)	Retained reads (millions)	BAM records (millions)	Catalogue loci
Coral ddRAD	30	213	205	142	285,987
Coral EpiRAD	30	156	149	102	164,411
Whitefish ddRAD	43	524	118	40	355,491
Whitefish EpiRAD	43	554	227	120	321,324

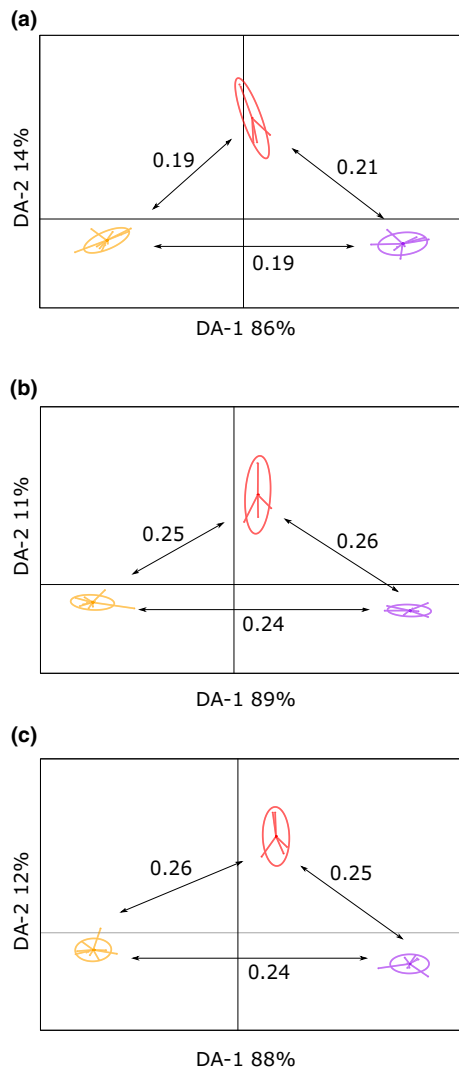
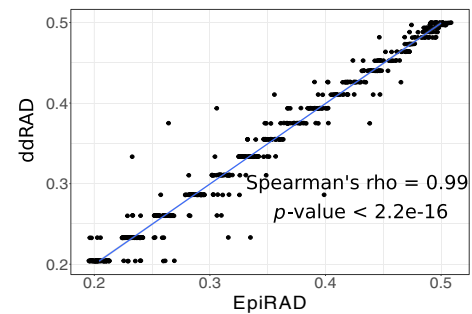


FIGURE 2 Results of the coral discriminant analysis of principal components (DAPC) analyses of the (a) single-nucleotide polymorphisms (SNPs) used by Dimond et al. (2017), (b) SNPs from the re-called ddRADseq dataset and (c) SNPs from the epiRADseq dataset. The analysis was based on five retained principal components, as suggested by the cross-validation of DAPC. These PCs were then summarized with two discriminant functions and percent variance captured appears on the axes. The numbers on arrows are Weir and Cockerham F_{ST} values between the clusters

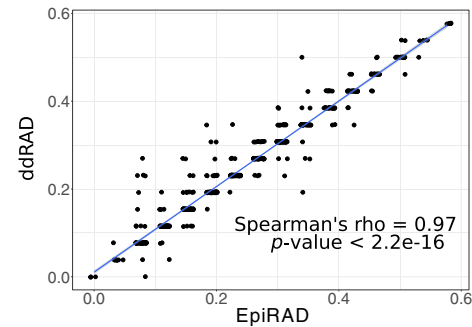
the ddRADseq library retained 118 M reads, while the epiRADseq library retained 227 M reads. After mapping to the reference genome, the ddRADseq library retained 40 M reads, while the epiRADseq library retained 120 M reads (Table 1). Excluding the samples with fewer than 350 K reads left a total of 23 samples plus two technical replicates in the epiRADseq dataset and 23 samples plus two technical replicates in the ddRADseq dataset.

The *Stacks* pipeline produced a catalogue of 355,491 loci for the ddRADseq library, with a mean effective per sample coverage of 12.7 \times , and of 321,324 loci for the epiRADseq library, with a mean effective per sample coverage of 35 \times . The average number of loci

(a) Expected heterozygosity



(b) Observed heterozygosity



(c) Nucleotide diversity

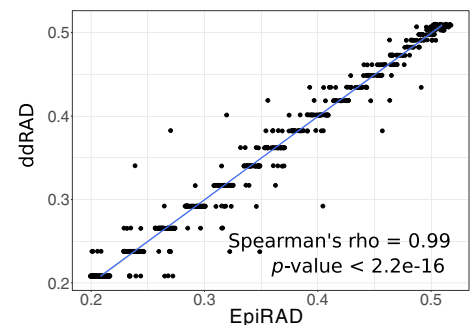


FIGURE 3 Correlation of (a) expected heterozygosity, (b) observed heterozygosity and (c) nucleotide diversity between ddRADseq (y axis) and epiRADseq (x axis) estimates for the coral data. Each dot represents a genomic site from the 'sumstats.tsv' file of the *Stacks* pipeline that was shared between the ddRADseq and the epiRADseq datasets

per individual was 108,127 and 110,614 for the ddRADseq and epiRADseq respectively.

3.4 | Whitefish population genetic data analysis

The number of SNPs retained was very similar for those generated with the epiRADseq method and the ddRADseq method and decreased with increasing filtering stringency (Figure 1b); for the epiRADseq data, we recovered 6,999, 6,699 and 5,559 SNPs in the $-r$ 67, $-r$ 75 and $-r$ 100 datasets, respectively, while for the ddRADseq data, we recovered

7,264, 6,969 and 5,269 SNPs in the three datasets respectively. A total of 4,538 SNPs were shared between the two $-r$ 67 datasets, 4,313 SNPs were shared between the two $-r$ 75 datasets and 2,987 SNPs were shared between the two $-r$ 100 datasets.

The results of the population genetic structure analysis with DAPC were consistent across filtering stringencies and datasets (Figure 4). The four populations grouped into two genetic clusters, with the two translocated populations clustering with their source population, respectively, and separating on axis 1 (and so displayed on one axis of variation instead of the two axes shown for the corals). F_{ST} divergence between the two clusters was identical between methods for the $-r$ 67 and $-r$ 75 datasets at $F_{ST} = 0.23$, and it was slightly higher for the ddRADseq in the $-r$ 100 datasets at 0.24 and 0.25 (Figure 4).

The estimates of heterozygosity and nucleotide diversity inferred from ddRADseq- and epiRADseq-derived SNPs were highly correlated, with Spearman's correlations of 88.5%–92.8% (Table 2).

3.5 | Genotyping error rate

The SNP genotyping error rate in the whitefish dataset was lower for epiRADseq for both analysis approaches. The *dist.gene* approach

recovered a mean error rate of 6% (\pm SD 0.6%) for the ddRADseq, and of 3% (\pm 0.5%) for the epiRADseq, while the Mastretta-Yanes et al. approach estimated a mean error of 5% (\pm 0.3%) for the ddRADseq and of 3% (\pm 0.4%) for the epiRADseq.

3.6 | Allele dropout

The coral data had a total of 102,647 observations and the whitefish data a total of 116,459 observations. The number of observations for which an individual was homozygous in the epiRADseq dataset but heterozygous in the ddRADseq dataset was 3,432 (epiRADseq-specific ADO, 3.3% of all shared loci in the data on average per individual) for the coral data, and 5,989 (epiRADseq-specific ADO, 5.1% of loci) for the whitefish data. The number of observations for which an individual was heterozygous in the epiRADseq dataset but homozygous in the ddRADseq dataset was 2,604 (ddRADseq-specific ADO, 2.5%) for the coral data and 4,533 (ddRADseq-specific ADO, 3.9%) for the whitefish data. The difference and therefore the ADO that is epiRADseq-specific, assuming an equal level of technical ADO for epiRADseq and ddRADseq (i.e. ddRADseq-specific) ADO, is 0.8% in coral and 1.2% in whitefish data.

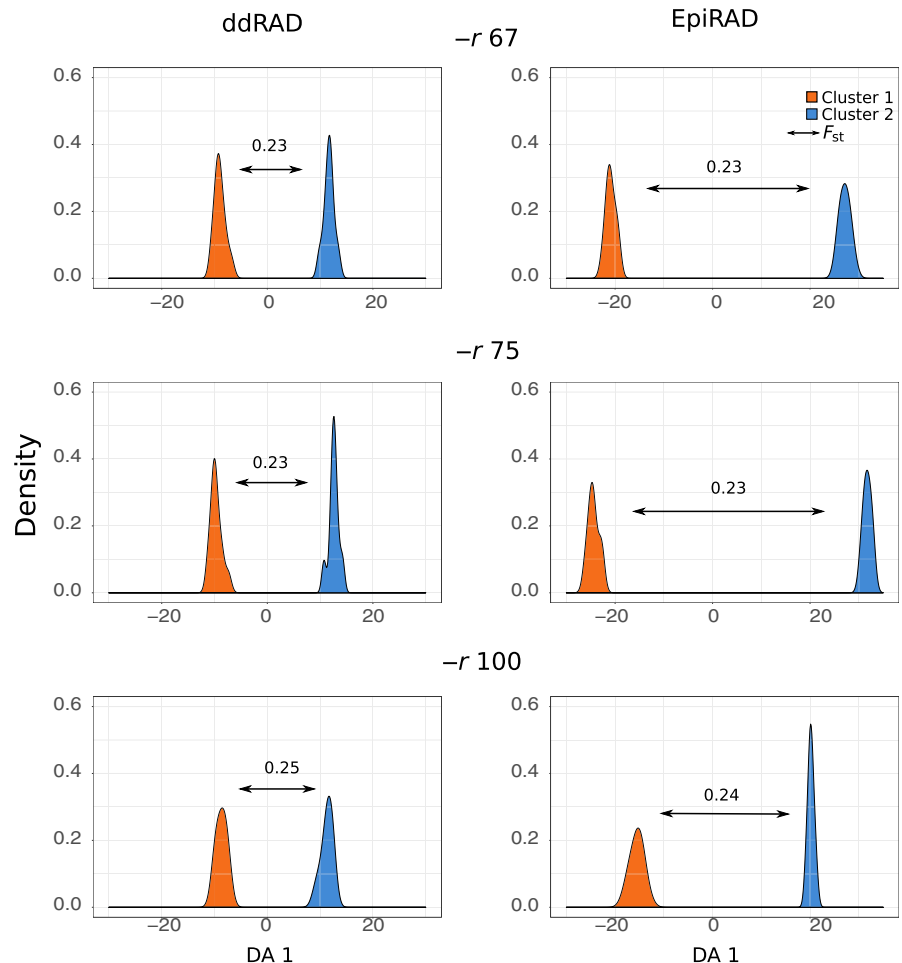


FIGURE 4 Results of the European whitefish discriminant analysis of principal components (DAPC) analyses at three different filtering stringencies ($-r$ 67, $-r$ 75 and $-r$ 100). The analysis was based on five retained principal components, as suggested by the cross-validation of DAPC. These PCs were then summarized on one discriminant function, as only two genetic clusters are observed. The numbers above arrows represent Weir and Cockerham F_{ST} values between the two identified clusters

TABLE 2 Spearman's correlation between coral and whitefish epiRADseq and ddRADseq estimates of expected heterozygosity (H_e), observed heterozygosity (H_o) and nucleotide diversity (π) for $-r$ 67, $-r$ 75 and $-r$ 100 datasets. Number of sites corresponds to the single-nucleotide polymorphisms shared between epiRADseq and ddRADseq datasets, for which the correlation was calculated

	Stacks filtering	Number of sites	H_e	H_o	π
Whitefish	$-r$ 67	4,538	0.904	0.884	0.896
	$-r$ 75	4,313	0.911	0.888	0.903
	$-r$ 100	2,987	0.928	0.906	0.919
Coral	$-r$ 100	676	0.988	0.972	0.988

4 | DISCUSSION

Here we used two independent natural animal populations datasets to show that epiRADseq data can be used to derive SNPs for population genomic analyses. We compared SNP number, estimates of summary statistics and inference of population structure between ddRADseq and epiRADseq methods in a newly generated dataset of European whitefish and a previously published dataset on corals, allowing us to demonstrate the robustness of the molecular methods and of the bioinformatic pipelines independently. Overall, we found strong agreement for all of the above metrics between epiRADseq and ddRADseq protocols, meaning that epiRADseq data give equivalent results to the well-established method of ddRADseq-derived SNPs. The implication is that a single dataset can be used for epigenetic analyses and for inference of population structure. This is not only efficient but valuable for studies on the association between epigenetic and genetic diversity and their relationship with phenotype.

4.1 | SNPs from epiRADseq and population genetics: Coral data

The coral dataset was drawn from Dimond et al. (2017), where they investigated population structure between three morphospecies of coral with ddRADseq and looked at the relationship between DNA methylation and environmental factors. The number of SNPs we recovered for our dataset is slightly lower than those used in Dimond et al. (2017); we retained 1,046 SNPs while they retained 1,113 SNPs. This minor difference is likely because different bioinformatic pipelines were applied, as they used *PyRAD* (Eaton, 2014) while we used *Stacks* (Catchen et al., 2013). We newly recovered 819 SNPs from their epiRADseq data.

Our genetic diversity, differentiation and population structure results of the coral data, derived from SNPs from their epiRADseq data, are consistent with those obtained by Dimond et al. (2017) from ddRADseq. The F_{ST} estimates between the three population genetic clusters are slightly higher in our study, which is likely caused by the different loci retained by the *Stacks* versus *PyRAD* pipelines

(Pante et al., 2015). However, F_{ST} results are rarely strictly comparable across studies and instead are relative to the markers used (Hartl & Clark, 2007), and these deviations can be considered irrelevant. These explorations and comparisons of our pipeline on the coral dataset demonstrate the appropriateness of the pipelines we applied and that the baseline genetic information is comparable across studies.

4.2 | SNPs from epiRADseq and population genetics: Fish data

We explored the effect of different filtering levels on the SNP retention of epiRADseq and ddRADseq for the whitefish data. We did not explore this with the coral data because we were interested in comparing population structure estimates between epiRADseq and the published estimates derived from ddRADseq (per Dimond et al., 2017). As expected, the $-r$ 67 and $-r$ 75 ddRADseq datasets had more SNPs than the respective epiRADseq datasets, but the epiRADseq $-r$ 100 dataset had more SNPs than the ddRADseq $-r$ 100 dataset. This is probably due to the higher coverage of the epiRADseq reads (85 M reads for 25 individuals in the epiRADseq versus 32 M reads for 25 individuals in the ddRADseq), which resulted in more SNPs retained in the most stringently filtered dataset. This is library-specific and will need to be assessed in each experiment.

We find an agreement between ddRADseq and epiRADseq analyses of population structure in the whitefish data, as both methods recover two clusters in our dataset of four sampled and closely related populations. The $-r$ filtering had some impact on the correlation of the summary statistics between ddRADseq and epiRADseq, with the correlation increasing from as low as 88% up to 92% as the filtering became more stringent. However, with regard to the use of SNPs from epiRADseq, it is important to consider that comparability across different datasets is not what matters; here that is done to evidence the method. Each of these stringencies and datasets could be valid in an empirical study. Overall, these results suggest that allowing some missing data (i.e. $-r$ of 67% or 75%) will not bias genetic analyses conducted with SNPs from epiRADseq data, consistent with what has already been shown previously with ddRADseq (Shafer et al., 2017).

4.3 | Evaluating the robustness of SNPs from epiRADseq

Genotyping error in NGS techniques is due to several factors, including sequencing errors, assembly errors and missing data and will be influenced by coverage (Mastretta-Yanes et al., 2015). Using technical replicates is a way to estimate this error, which can then be moderated by fine-tuning the bioinformatic pipeline. We found that the SNP genotyping error rate is low and very similar between ddRADseq and epiRADseq libraries, ranging between 3% and 6% according to the calculation method used. Mastretta-Yanes et al. (2015) found SNP error rates between 2.4% and 5.8% using the *Stacks* pipeline

on Illumina-based RAD sequencing. Recknagel et al. (2015), using a similar laboratory protocol as used for the whitefish libraries here but sequenced on an Ion Proton platform, recovered genotyping errors of 1.8%–2.2%. Dimond et al. (2017) used the ddRADseq and epiRADseq samples as technical replicates, as they were sequenced on the same lanes, and recovered a mean genotyping error rate of 3.6% (SD 3.1%). Therefore, genotyping error rates in the whitefish libraries are consistent with those found by previous studies and are very similar between the ddRADseq and epiRADseq approaches.

We distinguish possible causes of missing data in SNPs from epiRADseq and assess their potential effects on genetic estimates. The first is a possible lower number of loci and SNPs recovered from epiRADseq libraries (compared to ddRADseq) as seen in the fish and to a greater extent, the coral datasets (Figure 1). Methylation of the cut site on both DNA strands can explain this, as it prevents loci being sequenced. However, the reduction in loci number is not expected to have an effect on population genetic estimates when best practices for data filtering and loci sharing across populations are followed (Paris, Stevens, & Catchen, 2017; Shafer et al., 2017). Second, we assessed cases in epiRADseq where individuals are incorrectly assigned a homozygous genotype at a site because one allele is methylated (therefore not sequenced) while the other allele is not methylated and therefore sequenced. This methylation-induced ADO is a biological factor of the data that will be particular to epiRADseq (Schield et al., 2016). Third, ADO could be due to technical factors such as library preparation, PCR and/or sequencing artefacts that result in incomplete coverage of all alleles at all loci. This effect is common to all genotyping-by-sequencing (Gautier et al., 2013) and not particular to epiRADseq. Here we calculated the contribution of this technical effect as ddRADseq-specific ADO.

CpG methylation is mostly symmetrical in animals and occurs on both strands (Selker, Fritz, & Singer, 1993). Therefore, the amount of epiRADseq-specific ADO is expected to be low when using a restriction enzyme targeting CpG methylation, as the locus will not be sequenced. In both coral and whitefish data, we found evidence of a small amount of epiRADseq-specific ADO, between 0.8% (coral) and 1.2% (whitefish) of all loci per individual. This agrees with previous studies that found half-methylation (i.e. methylation on one strand only) to be <2% at CpG sites (Bird, 1978), indicating that the ADO bias introduced by epiRADseq for genetic analyses is exceedingly low.

In our results, this is also reflected by comparable estimates of heterozygosity and F_{ST} between populations obtained from ddRADseq and epiRADseq SNPs. However, the amount of ADO may vary by species so only future research will show how reflective these two exemplars—which are distantly related species of different phyla, with different methylation landscapes and different levels of genetic variation (Metzger & Schulte, 2016)—are of other species.

Assessing the coral and whitefish data together, we find agreement for population structure estimates either with ddRADseq or epiRADseq. However, the percentage of SNPs shared between ddRADseq and epiRADseq was higher in the coral data (83% vs. 55%–65%). This could

be explained by genome complexity and size differences between the two organisms. Salmonids, including whitefish, have undergone a recent whole-genome duplication (Macqueen & Johnston, 2014) and have a genome size of ~3 Gb (Gregory, 2018). Scleractinian corals, such as *Porites* spp., have genomes ranging from 420 Mb to 1.14 Gb (Gregory, 2018). Smaller genomes generate fewer RAD loci for a given enzyme, which are then more likely to be found across sequencing libraries at a given coverage (see Recknagel et al., 2015 for detailed quantifications). Furthermore, DNA methylation levels and patterns differ between the organisms studied here and may have an impact. Most of the CpG sites (~80%) in vertebrate genomes are methylated, with the unmethylated sites forming regions known as CpG islands, usually located near gene promoters (Metzger & Schulte, 2016). In contrast, most of the methylation in invertebrates occurs specifically in CpG sites within gene bodies (Liew et al., 2018). The methylation level of CpG sites in the scleractinian coral *Stylophora pistillata* is around 7% (Liew et al., 2018), a stark contrast to the methylation level of vertebrates. Interspecific methylation differences might influence the number of fragments cut during digestion with *HpaII* and therefore affect the number of loci sequenced. We did not explore the genomic location of the SNPs used here, but, with appropriate reference genome annotation information, that is possible and would be very informative. Overall, further research is needed on lineage- or population-specific methylation and how it might affect SNP recovery.

4.4 | Recommendations

We propose some recommendations for researchers interested in conducting epigenomic and genomic analyses on the same individuals/populations using epiRADseq and SNPs derived from it. Many of these recommendations are shared with best practice for genotyping-by-sequencing of any type, because all have stochastic variations and potentially biasing and non-biasing artefacts (see e.g. O'Leary, Puritz, Willis, Hollenbeck, & Portnoy, 2018; Paris et al., 2017; Recknagel et al., 2015; Shafer et al., 2017).

First, SNPs from epiRADseq have only been assessed in the species presented here (an invertebrate [coral] and a vertebrate [whitefish]), so it is not known if the negligible level of epiRADseq-specific ADO we found is generalizable. Therefore, for most cautious inference, we recommend a preliminary analysis to quantify epiRADseq-specific ADO and optimize library specifications by conducting ddRADseq and epiRADseq together on a subset of individuals. Following the pipelines for comparison presented here, researchers can then determine the level of ddRADseq-specific and epiRADseq-specific ADO, the number of loci retained and the consistency of the datasets across population genetic analyses to be applied. If results are sufficiently similar between the SNP datasets derived from ddRADseq and epiRADseq, the remaining samples can be sequenced only with epiRADseq, thereby halving the subsequent library and sequencing costs. We note that the level of 'sufficient' similarity between the datasets would need to be evaluated on a case-by-case basis depending on the study system, the research

question and the sensitivity of the population genetic analyses to ADO (see Gautier et al., 2013).

Our other recommendations are also relevant to genotyping-by-sequencing. For example, technical replicates provide an estimate of genotyping error rate (Mastretta-Yanes et al., 2015); sufficient sequencing coverage per locus per individual may need to be empirically determined (Paris et al., 2017; Recknagel et al., 2015); bioinformatic pipelines need to be evaluated for each dataset (for extensive treatment see, e.g., O'Leary et al., 2018; Paris et al., 2017; Shafer et al., 2017). If ddRADseq SNPs are not present to validate epiRADseq loci for differential methylation analysis (following the Dimond et al., 2017 approach), future studies could test a filtering strategy aimed at removing epiRADseq loci that have zero reads in one or more individuals/populations/experimental treatments in case these are lost due to genotype variation rather than methylation. This would ensure that only informative loci are retained in the analysis. See Schield et al. (2016) for details on the evaluation of epiRADseq library preparation, sequencing and data for methylation analysis.

5 | CONCLUSIONS

Here, we showed that the recently developed epiRADseq approach for the study of DNA methylation variation was also suitable for generating SNPs for population genetic analyses, using both reference-based and de novo approaches. Sequencing only an epiRADseq library halves the cost in time, consumables and sequencing compared to doing ddRADseq for SNPs and epiRADseq for methylation abundance. This combination provides informative biological data for population genomics and differential methylation, which is a topic of growing interest in molecular ecology and evolution for its heritable and non-heritable effects.

ACKNOWLEDGEMENTS

This work was funded by the University of Glasgow College of Medical, Veterinary and Life Sciences doctoral training programme. We thank J.L. Dimond, S.K. Gamblewood and S.B. Roberts for making their data public, so we could explore it for this paper. We thank Glasgow Polyomics and J. Galbraith for sequencing; M. Capstick for support in the laboratory; A. Jacobs and H. Recknagel for analysis advice; and A. Jacobs, A. Crespel and four anonymous reviewers for comments on the draft manuscript. For access to unpublished lake whitefish scaffolds, we thank L. Bernatchez, C. Rougeux, S. Pavey, E. Normandeau, S. Lien and T. Nome. We declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

K.R.E. conceived the idea and designed the overall methodology with input from M.C.; M.C., K.R.E. and C.E.A. collected whitefish samples; M.C. generated and analysed all the data; M.C. led the writing of the manuscript with input from K.R.E. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Data and scripts are archived and made available in University of Glasgow Enlighten Repository <https://doi.org/10.5525/gla.researchdata.963> (Crotti, Adams, & Elmer, 2020).

ORCID

Marco Crotti  <https://orcid.org/0000-0002-8619-7988>

Colin E. Adams  <https://orcid.org/0000-0003-2470-9754>

Kathryn R. Elmer  <https://orcid.org/0000-0002-9219-7001>

REFERENCES

- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Bird, A. P. (1978). Use of restriction enzymes to study eukaryotic DNA methylation. II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *Journal of Molecular Biology*, 118(1), 49–60. [https://doi.org/10.1016/0022-2836\(78\)90243-7](https://doi.org/10.1016/0022-2836(78)90243-7)
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bosssdorf, O., Richards, C. L., & Pigliucci, M. (2007). Epigenetics for ecologists. *Ecology Letters*. <https://doi.org/10.1111/j.1461-0248.2007.01130.x>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chong, Z., Ruan, J., & Wu, C.-I. (2012). Rainbow: An integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, 28(21), 2732–2737. <https://doi.org/10.1093/bioinformatics/bts482>
- Crotti, M., Adams, C. E., & Elmer, K. R. (2020). Data from: Population genomic SNPs from epigenetic RADs: Gaining genetic and epigenetic data from a single established next-generation sequencing approach. *Enlighten Repository*. <https://doi.org/10.5525/gla.researchdata.963>
- Dimond, J. L., Gamblewood, S. K., & Roberts, S. B. (2017). Genetic and epigenetic insight into morphospecies in a reef coral. *Molecular Ecology*, 26(19), 5031–5042. <https://doi.org/10.1111/mec.14252>
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: Emerging patterns and implications. *Nature Reviews Genetics*, 13(2), 97–109. <https://doi.org/10.1038/nrg3142>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdellhué, C., Pudlo, P., ... Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22(11), 3165–3178. <https://doi.org/10.1111/mec.12089>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Gregory, T. R. (2018). *Animal genome size database*. Retrieved from <http://www.genomesize.com/index.php>
- Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics*. Sunderland, MA: Sinauer Associates.
- Hu, J., & Barrett, R. D. H. (2017). Epigenetics in natural animal populations. *Journal of Evolutionary Biology*, 30(9), 1612–1632. <https://doi.org/10.1111/jeb.13130>

- Jacobs, A., Hughes, M., Robinson, P., Adams, C., & Elmer, K. (2018). The genetic architecture underlying the evolution of a rare piscivorous life history form in brown trout after secondary contact and strong introgression. *Genes*, 9(6), 280. <https://doi.org/10.3390/genes9060280>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Klughammer, J., Datlinger, P., Printz, D., Sheffield, N. C., Farlik, M., Hadler, J., ... Bock, C. (2015). Differential DNA methylation analysis without a reference genome. *Cell Reports*, 13(11), 2621–2633. <https://doi.org/10.1016/j.celrep.2015.11.024>
- Leontiou, C. A., Hadjidaniel, M. D., Mina, P., Antoniou, P., Ioannides, M., & Patsalis, P. C. (2015). Bisulfite conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. *PLoS ONE*, 10(8), e0135058. <https://doi.org/10.1371/journal.pone.0135058>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liew, Y. J., Zoccola, D., Li, Y., Tambutté, E., Venn, A. A., Michell, C. T., ... Aranda, M. (2018). Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *Science Advances*, 4(6), eaar8028. <https://doi.org/10.1126/sciadv.aar8028>
- Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778), 20132881. <https://doi.org/10.1098/rspb.2013.2881>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. <https://doi.org/10.1111/1755-0998.12291>
- Metzger, D. C. H., & Schulte, P. M. (2016). Epigenomics in marine fishes. *Marine Genomics*, 30, 43–54. <https://doi.org/10.1016/J.MARGEN.2016.01.004>
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27(16), 3193–3206. <https://doi.org/10.1111/mec.14792>
- Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M. C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, 114(5), 450–459. <https://doi.org/10.1038/hdy.2014.105>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, 8(10), 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Recknagel, H., Jacobs, A., Herzyk, P., & Elmer, K. R. (2015). Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRAD-seq-ion) with nonmodel organisms. *Molecular Ecology Resources*, 15(6), 1316–1329. <https://doi.org/10.1111/1755-0998.12406>
- Schild, D. R., Walsh, M. R., Card, D. C., Andrew, A. L., Adams, R. H., & Castoe, T. A. (2016). EpiRADseq: Scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods in Ecology and Evolution*, 7(1), 60–69. <https://doi.org/10.1111/2041-210X.12435>
- Schrey, A. W., Alvarez, M., Foust, C. M., Kilvitis, H. J., Lee, J. D., Liebl, A. L., ... Robertson, M. (2013). Ecological epigenetics: Beyond MS-AFLP. *Integrative and Comparative Biology*, 53(2), 340–350. <https://doi.org/10.1093/icb/ict012>
- Selker, E. U., Fritz, D. Y., & Singer, M. J. (1993). Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. *Science*, 262(5140), 1724–1728. <https://doi.org/10.1126/science.8259516>
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917. <https://doi.org/10.1111/2041-210X.12700>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370.

How to cite this article: Crotti M, Adams CE, Elmer KR.

Population genomic SNPs from epigenetic RADs: Gaining genetic and epigenetic data from a single established next-generation sequencing approach. *Methods Ecol Evol*.

2020;11:839–849. <https://doi.org/10.1111/2041-210X.13395>